

OM protein - protein search, using sw model

Run on: April 21, 2004, 16:55:02 ; Search time 41 Seconds  
(without alignments)  
67.433 Million cell updates/sec

Title: US-10-057-789-41\_COPY\_1\_10  
Perfect score: 56  
Sequence: 1 CASENLYFQG 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1133595 seqs, 276475211 residues

Total number of hits satisfying chosen parameters: 1133595

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

%

Result	Query					
No.	Score	Match	Length	DB	ID	Description

1	56	100.0	11	14	US-10-057-789-41	Sequence 41, Appl
2	56	100.0	11	14	US-10-057-789-42	Sequence 42, Appl
3	56	100.0	11	14	US-10-212-628-41	Sequence 41, Appl
4	56	100.0	11	14	US-10-212-628-42	Sequence 42, Appl
5	56	100.0	12	14	US-10-057-789-43	Sequence 43, Appl
6	56	100.0	12	14	US-10-057-789-44	Sequence 44, Appl
7	56	100.0	12	14	US-10-212-628-43	Sequence 43, Appl
8	56	100.0	12	14	US-10-212-628-44	Sequence 44, Appl
9	47	83.9	18	14	US-10-057-789-7	Sequence 7, Appli
10	47	83.9	18	14	US-10-057-789-27	Sequence 27, Appl
11	47	83.9	18	14	US-10-057-789-30	Sequence 30, Appl
12	47	83.9	18	14	US-10-212-628-7	Sequence 7, Appli
13	47	83.9	18	14	US-10-212-628-27	Sequence 27, Appl
14	47	83.9	18	14	US-10-212-628-30	Sequence 30, Appl
15	47	83.9	19	14	US-10-057-789-2	Sequence 2, Appli
16	47	83.9	19	14	US-10-057-789-5	Sequence 5, Appli
17	47	83.9	19	14	US-10-057-789-8	Sequence 8, Appli
18	47	83.9	19	14	US-10-057-789-9	Sequence 9, Appli
19	47	83.9	19	14	US-10-057-789-10	Sequence 10, Appl
20	47	83.9	19	14	US-10-057-789-11	Sequence 11, Appl
21	47	83.9	19	14	US-10-057-789-12	Sequence 12, Appl
22	47	83.9	19	14	US-10-057-789-15	Sequence 15, Appl
23	47	83.9	19	14	US-10-057-789-17	Sequence 17, Appl
24	47	83.9	19	14	US-10-057-789-20	Sequence 20, Appl
25	47	83.9	19	14	US-10-057-789-22	Sequence 22, Appl
26	47	83.9	19	14	US-10-057-789-25	Sequence 25, Appl
27	47	83.9	19	14	US-10-057-789-28	Sequence 28, Appl
28	47	83.9	19	14	US-10-057-789-29	Sequence 29, Appl
29	47	83.9	19	14	US-10-057-789-31	Sequence 31, Appl
30	47	83.9	19	14	US-10-057-789-39	Sequence 39, Appl
31	47	83.9	19	14	US-10-057-789-40	Sequence 40, Appl
32	47	83.9	19	14	US-10-212-628-2	Sequence 2, Appli
33	47	83.9	19	14	US-10-212-628-5	Sequence 5, Appli
34	47	83.9	19	14	US-10-212-628-8	Sequence 8, Appli
35	47	83.9	19	14	US-10-212-628-9	Sequence 9, Appli
36	47	83.9	19	14	US-10-212-628-10	Sequence 10, Appl
37	47	83.9	19	14	US-10-212-628-11	Sequence 11, Appl
38	47	83.9	19	14	US-10-212-628-12	Sequence 12, Appl
39	47	83.9	19	14	US-10-212-628-15	Sequence 15, Appl
40	47	83.9	19	14	US-10-212-628-17	Sequence 17, Appl
41	47	83.9	19	14	US-10-212-628-20	Sequence 20, Appl
42	47	83.9	19	14	US-10-212-628-22	Sequence 22, Appl
43	47	83.9	19	14	US-10-212-628-25	Sequence 25, Appl
44	47	83.9	19	14	US-10-212-628-28	Sequence 28, Appl
45	47	83.9	19	14	US-10-212-628-29	Sequence 29, Appl

#### ALIGNMENTS

RESULT 1  
 US-10-057-789-41  
 ; Sequence 41, Application US/10057789  
 ; Publication No. US20030082522A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Paul Haynes

```

; APPLICANT: Jing Wei
; APPLICANT: John Yates
; APPLICANT: Nancy Andon
; TITLE OF INVENTION: DIFFERENTIAL LABELING FOR QUANTITATIVE
; TITLE OF INVENTION: ANALYSIS OF COMPLEX PROTEIN MIXTURES
; FILE REFERENCE: NADII.022A
; CURRENT APPLICATION NUMBER: US/10/057,789
; CURRENT FILING DATE: 2002-06-28
; PRIOR APPLICATION NUMBER: US 60/264,576
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/305,232
; PRIOR FILING DATE: 2001-07-13
; NUMBER OF SEQ ID NOS: 311
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 41
; LENGTH: 11
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthesized Peptide
US-10-057-789-41

```

```

Query Match          100.0%; Score 56; DB 14; Length 11;
Best Local Similarity 100.0%; Pred. No. 0.001;
Matches    10; Conservative    0; Mismatches    0; Indels    0; Gaps    0;

```

```

Qy      1 CASENLYFQG 10
        |||||
Db      1 CASENLYFQG 10

```

```

RESULT 3
US-10-212-628-41
; Sequence 41, Application US/10212628
; Publication No. US20030087329A1
; GENERAL INFORMATION:
; APPLICANT: Paul Haynes
; APPLICANT: Jing Wei
; APPLICANT: John Yates
; APPLICANT: Nancy Andon
; TITLE OF INVENTION: DIFFERENTIAL LABELING FOR QUANTITATIVE
; TITLE OF INVENTION: ANALYSIS OF COMPLEX PROTEIN MIXTURES
; FILE REFERENCE: NADII.022CP1
; CURRENT APPLICATION NUMBER: US/10/212,628
; CURRENT FILING DATE: 2002-08-01
; PRIOR APPLICATION NUMBER: US 60/264,576
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/305,232
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 10/057,789
; PRIOR FILING DATE: 2002-01-25
; NUMBER OF SEQ ID NOS: 311
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 41
; LENGTH: 11
; TYPE: PRT
; ORGANISM: Artificial Sequence

```

; FEATURE:  
; OTHER INFORMATION: Synthesized Peptide  
US-10-212-628-41

Query Match 100.0%; Score 56; DB 14; Length 11;  
Best Local Similarity 100.0%; Pred. No. 0.001;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CASENLYFQG 10  
| | | | | | | | | |  
Db 1 CASENLYFQG 10

Search completed: April 21, 2004, 16:59:51  
Job time : 41 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 21, 2004, 16:55:02 ; Search time 41 Seconds  
(without alignments)  
67.433 Million cell updates/sec

Title: US-10-057-789-41\_COPY\_1\_10  
Perfect score: 56  
Sequence: 1 CASENLYFQG 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1133595 seqs, 276475211 residues

Total number of hits satisfying chosen parameters: 1133595

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*

13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*  
 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*  
 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*  
 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	8		Query		DB	ID	Description
	Score	Match	Length				
1	56	100.0	11	14	US-10-057-789-41	Sequence 41, Appl	
2	56	100.0	11	14	US-10-057-789-42	Sequence 42, Appl	
3	56	100.0	11	14	US-10-212-628-41	Sequence 41, Appl	
4	56	100.0	11	14	US-10-212-628-42	Sequence 42, Appl	
5	56	100.0	12	14	US-10-057-789-43	Sequence 43, Appl	
6	56	100.0	12	14	US-10-057-789-44	Sequence 44, Appl	
7	56	100.0	12	14	US-10-212-628-43	Sequence 43, Appl	
8	56	100.0	12	14	US-10-212-628-44	Sequence 44, Appl	
9	47	83.9	18	14	US-10-057-789-7	Sequence 7, Appli	
10	47	83.9	18	14	US-10-057-789-27	Sequence 27, Appl	
11	47	83.9	18	14	US-10-057-789-30	Sequence 30, Appl	
12	47	83.9	18	14	US-10-212-628-7	Sequence 7, Appli	
13	47	83.9	18	14	US-10-212-628-27	Sequence 27, Appl	
14	47	83.9	18	14	US-10-212-628-30	Sequence 30, Appl	
15	47	83.9	19	14	US-10-057-789-2	Sequence 2, Appli	
16	47	83.9	19	14	US-10-057-789-5	Sequence 5, Appli	
17	47	83.9	19	14	US-10-057-789-8	Sequence 8, Appli	
18	47	83.9	19	14	US-10-057-789-9	Sequence 9, Appli	
19	47	83.9	19	14	US-10-057-789-10	Sequence 10, Appl	
20	47	83.9	19	14	US-10-057-789-11	Sequence 11, Appl	
21	47	83.9	19	14	US-10-057-789-12	Sequence 12, Appl	
22	47	83.9	19	14	US-10-057-789-15	Sequence 15, Appl	
23	47	83.9	19	14	US-10-057-789-17	Sequence 17, Appl	
24	47	83.9	19	14	US-10-057-789-20	Sequence 20, Appl	
25	47	83.9	19	14	US-10-057-789-22	Sequence 22, Appl	
26	47	83.9	19	14	US-10-057-789-25	Sequence 25, Appl	
27	47	83.9	19	14	US-10-057-789-28	Sequence 28, Appl	
28	47	83.9	19	14	US-10-057-789-29	Sequence 29, Appl	
29	47	83.9	19	14	US-10-057-789-31	Sequence 31, Appl	
30	47	83.9	19	14	US-10-057-789-39	Sequence 39, Appl	
31	47	83.9	19	14	US-10-057-789-40	Sequence 40, Appl	
32	47	83.9	19	14	US-10-212-628-2	Sequence 2, Appli	
33	47	83.9	19	14	US-10-212-628-5	Sequence 5, Appli	
34	47	83.9	19	14	US-10-212-628-8	Sequence 8, Appli	
35	47	83.9	19	14	US-10-212-628-9	Sequence 9, Appli	
36	47	83.9	19	14	US-10-212-628-10	Sequence 10, Appl	
37	47	83.9	19	14	US-10-212-628-11	Sequence 11, Appl	
38	47	83.9	19	14	US-10-212-628-12	Sequence 12, Appl	
39	47	83.9	19	14	US-10-212-628-15	Sequence 15, Appl	
40	47	83.9	19	14	US-10-212-628-17	Sequence 17, Appl	
41	47	83.9	19	14	US-10-212-628-20	Sequence 20, Appl	

42	47	83.9	19	14	US-10-212-628-22	Sequence 22, Appl
43	47	83.9	19	14	US-10-212-628-25	Sequence 25, Appl
44	47	83.9	19	14	US-10-212-628-28	Sequence 28, Appl
45	47	83.9	19	14	US-10-212-628-29	Sequence 29, Appl

#### ALIGNMENTS

Search completed: April 21, 2004, 16:59:51  
Job time : 41 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 21, 2004, 16:40:05 ; Search time 23 Seconds  
(without alignments)  
22.446 Million cell updates/sec

Title: US-10-057-789-41\_COPY\_1\_10  
Perfect score: 56  
Sequence: 1 CASENLYFQG 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	43	76.8	342	3	US-08-828-741B-6	Sequence 6, Appli
2	43	76.8	342	4	US-09-160-567-6	Sequence 6, Appli
3	43	76.8	342	4	US-09-710-299-6	Sequence 6, Appli
4	43	76.8	342	4	US-09-509-031-6	Sequence 6, Appli
5	43	76.8	482	4	US-09-509-031-16	Sequence 16, Appl
6	43	76.8	495	3	US-08-828-741B-4	Sequence 4, Appli
7	43	76.8	495	4	US-09-160-567-4	Sequence 4, Appli
8	43	76.8	495	4	US-09-710-299-4	Sequence 4, Appli
9	43	76.8	495	4	US-09-509-031-4	Sequence 4, Appli
10	40	71.4	25	4	US-09-039-780A-96	Sequence 96, Appl
11	40	71.4	37	4	US-09-039-780A-97	Sequence 97, Appl

12	40	71.4	245	4	US-09-280-030-66	Sequence 66, Appl
13	40	71.4	619	4	US-09-596-248D-59	Sequence 59, Appl
14	40	71.4	621	4	US-09-898-297-1	Sequence 1, Appli
15	40	71.4	621	4	US-09-995-099-1	Sequence 1, Appli
16	40	71.4	1063	4	US-09-596-248D-47	Sequence 47, Appl
17	39	69.6	7	1	US-08-021-603A-3	Sequence 3, Appli
18	39	69.6	7	1	US-08-021-603A-11	Sequence 11, Appl
19	39	69.6	7	2	US-08-846-021A-12	Sequence 12, Appl
20	39	69.6	7	5	PCT-US94-01176-3	Sequence 3, Appli
21	39	69.6	7	5	PCT-US94-01176-11	Sequence 11, Appl
22	39	69.6	12	1	US-08-021-603A-6	Sequence 6, Appli
23	39	69.6	12	5	PCT-US94-01176-6	Sequence 6, Appli
24	39	69.6	24	1	US-08-021-603A-16	Sequence 16, Appl
25	39	69.6	24	5	PCT-US94-01176-16	Sequence 16, Appl
26	36	64.3	48	3	US-09-177-249-202	Sequence 202, App
27	36	64.3	248	4	US-09-540-236-2035	Sequence 2035, Ap
28	35	62.5	506	4	US-09-107-532A-5363	Sequence 5363, Ap
29	34	60.7	13	4	US-09-280-030-2	Sequence 2, Appli
30	34	60.7	30	4	US-09-039-780A-98	Sequence 98, Appl
31	34	60.7	30	4	US-09-039-780A-100	Sequence 100, App
32	34	60.7	44	4	US-09-039-780A-99	Sequence 99, Appl
33	34	60.7	106	4	US-09-543-681A-4251	Sequence 4251, Ap
34	34	60.7	116	3	US-09-027-449-50	Sequence 50, Appl
35	34	60.7	116	3	US-08-804-444A-50	Sequence 50, Appl
36	34	60.7	116	3	US-09-026-985-50	Sequence 50, Appl
37	34	60.7	116	4	US-09-121-952A-50	Sequence 50, Appl
38	34	60.7	116	4	US-09-234-340A-50	Sequence 50, Appl
39	34	60.7	140	4	US-09-280-030-64	Sequence 64, Appl
40	34	60.7	322	2	US-08-622-354-3	Sequence 3, Appli
41	34	60.7	1034	4	US-09-252-991A-20969	Sequence 20969, A
42	33	58.9	7	3	US-08-782-480-38	Sequence 38, Appl
43	33	58.9	7	3	US-08-954-211-38	Sequence 38, Appl
44	33	58.9	7	4	US-09-005-167A-38	Sequence 38, Appl
45	33	58.9	7	4	US-09-176-741B-38	Sequence 38, Appl

#### ALIGNMENTS

#### RESULT 1

US-08-828-741B-6

; Sequence 6, Application US/08828741B

; Patent No. 6043069

; GENERAL INFORMATION:

; APPLICANT: Koentgen, Frank

; APPLICANT: Suess, Gabriele M.

; APPLICANT: Tarlinton, David M.

; APPLICANT: Treutlein, Herbert R.

; TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF

; TITLE OF INVENTION: PRODUCING SAME

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER

; STREET: 400 Garden City Plaza

; CITY: Garden City

; STATE: New York

; COUNTRY: United States of America



```

; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/828,741B
; FILING DATE: 26-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: DiGiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 10591
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 342 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-828-741B-6

```

```

Query Match          76.8%; Score 43; DB 3; Length 342;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

```

```

Qy          3 SENLYFQG 10
            |||||||
Db          165 SENLYFQG 172

```

```

RESULT 10
US-09-039-780A-96
; Sequence 96, Application US/09039780A
; Patent No. 6376248
; GENERAL INFORMATION:
; APPLICANT: HAWLEY-NELSON, PAMELA
;           LAN, JIANQING
;           SHIH, POJEN
;           JESSE, JOEL A.
;           SCHIFFERLI, KEVIN P.
;           GEBEYEHU, GULILAT
; TITLE OF INVENTION: PEPTIDE-ENHANCED TRANSFECTIONS
; NUMBER OF SEQUENCES: 120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GREENLEE, WINNER & SULLIVAN
; STREET: 5370 MANHATTAN CIRCLE, SUITE 201
; CITY: BOULDER
; STATE: CO
; COUNTRY: US
; ZIP: 80303

```

```

;      COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: PatentIn Release #1.0, Version #1.30
;      CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/09/039,780A
;      FILING DATE: 16-Mar-1998
;      CLASSIFICATION: <Unknown>
;      ATTORNEY/AGENT INFORMATION:
;      NAME: SULLIVAN, SALLY A.
;      REGISTRATION NUMBER: 32,064
;      REFERENCE/DOCKET NUMBER: 32-95C
;      TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (303)499-8080
;      TELEFAX: (303)499-8089
;      INFORMATION FOR SEQ ID NO: 96:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH: 25 amino acids
;      TYPE: amino acid
;      STRANDEDNESS: not relevant
;      TOPOLOGY: linear
;      MOLECULE TYPE: peptide
;      HYPOTHETICAL: NO
;      ANTI-SENSE: NO
;      SEQUENCE DESCRIPTION: SEQ ID NO: 96:
US-09-039-780A-96

```

```

Query Match          71.4%;  Score 40;  DB 4;  Length 25;
Best Local Similarity 87.5%;  Pred. No. 0.81;
Matches      7;  Conservative    1;  Mismatches    0;  Indels      0;  Gaps      0;

```

```

Qy      3 SENLYFQG 10
        :|||||||
Db      17 TENLYFQG 24

```

```

Search completed: April 21, 2004, 16:55:22
Job time : 32 secs

```

=> fil hcaplus  
FILE 'HCAPLUS' ENTERED AT 15:57:58 ON 21 APR 2004  
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.  
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.  
COPYRIGHT (C) 2004 AMERICAN CHEMICAL SOCIETY (ACS)

Copyright of the articles to which records in this database refer is held by the publishers listed in the PUBLISHER (PB) field (available for records published or updated in Chemical Abstracts after December 26, 1996), unless otherwise indicated in the original publications. The CA Lexicon is the copyrighted intellectual property of the the American Chemical Society and is provided to assist you in searching databases on STN. Any dissemination, distribution, copying, or storing of this information, without the prior written consent of CAS, is strictly prohibited.

FILE COVERS 1907 - 21 Apr 2004 VOL 140 ISS 17  
FILE LAST UPDATED: 20 Apr 2004 (20040420/ED)

This file contains CAS Registry Numbers for easy and accurate substance identification.

=>  
=>

=> d stat que  
L65 127 SEA FILE=REGISTRY ABB=ON PLU=ON CASEN/SQSP  
L66 79 SEA FILE=HCAPLUS ABB=ON PLU=ON L65  
L67 377 SEA FILE=REGISTRY ABB=ON PLU=ON LYFQG/SQSP  
L68 4 SEA FILE=REGISTRY ABB=ON PLU=ON L66 AND L67  
L69 1 SEA FILE=HCAPLUS ABB=ON PLU=ON L68

=>  
=>

=> d ibib abs hitrn l69

L69 ANSWER 1 OF 1 HCAPLUS COPYRIGHT 2004 ACS on STN  
ACCESSION NUMBER: 2002:575099 HCAPLUS  
DOCUMENT NUMBER: 137:137275  
TITLE: Differential labeling for quantitative analysis of complex protein mixtures  
INVENTOR(S): Haynes, Paul; Wei, Jing; Yates, John; Andon, Nancy  
PATENT ASSIGNEE(S): Syngenta Participation Ag, USA  
SOURCE: PCT Int. Appl., 79 pp.  
CODEN: PIXXD2  
DOCUMENT TYPE: Patent  
LANGUAGE: English  
FAMILY ACC. NUM. COUNT: 1  
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2002059144	A2	20020801	WO 2002-US2487	20020125
WO 2002059144	A3	20031218		

W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ,

UA, UG, UZ, VN, YU, ZA, ZM, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM  
 RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW, AT, BE, CH,  
 CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR,  
 BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG  
 US 2003082522 A1 20030501 US 2002-57789 20020125  
 US 2003087329 A1 20030508 US 2002-212628 20020801  
 WO 2004013636 A2 20040212 WO 2003-IB3863 20030728

W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN,  
 CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,  
 GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR,  
 LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH,  
 PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ,  
 UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW, AM, AZ, BY, KG, KZ, MD, RU,  
 TJ, TM  
 RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW, AT, BE, BG,  
 CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC,  
 NL, PT, RO, SE, SI, SK, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GQ,  
 GW, ML, MR, NE, SN, TD, TG

PRIORITY APPLN. INFO.: US 2001-264576P P 20010126  
 US 2001-305232P P 20010713  
 US 2002-57789 A1 20020125  
 US 2002-212628 A 20020801

OTHER SOURCE(S): MARPAT 137:137275

AB The invention concerns a method of simultaneously identifying and detg.  
 the levels of expression of cysteine-contg. proteins in normal and  
 perturbed cells, a method for proteomic anal., a process for prepg. fusion  
 proteins, and compds. and reagents related thereto. This invention  
 provides methods and reagents that can be employed in proteome anal. which  
 overcome the limitations inherent in traditional techniques The basic  
 approach described can be employed for the quant. anal. of protein  
 expression in complex samples (such as cells, tissues, and fractions  
 thereof), the detection and quantitation of specific proteins in complex  
 samples, and the quant. measurement of specific enzymic activities in  
 complex samples. We have designed trifunctional synthetic peptide based  
 reagents that can be used for reducing the complexity of peptide mixts. by  
 labeling peptides with iodoacetamido groups and then selectively enriching  
 only those peptides contg. labeled cysteine residues. Embodiments of this  
 invention provide anal. reagents and mass spectrometry-based methods using  
 these reagents for the rapid and quant. anal. of proteins or protein  
 function in mixts. of proteins. The anal. method can be used for qual.  
 and particularly for quant. anal. of global protein expression profiles in  
 cells and tissues, i.e., the quant. anal. of proteomes.

IT 444196-45-6DP, acyl derivs. 444196-46-7DP, acyl derivs.  
 444196-47-8DP, acyl derivs. 444196-48-9DP, acyl derivs.  
 RL: BSU (Biological study, unclassified); PRP (Properties); SPN (Synthetic  
 preparation); BIOL (Biological study); PREP (Preparation)  
 (differential labeling for quant. anal. of complex protein mixts.)

=>  
 =>

=> fil reg  
 FILE 'REGISTRY' ENTERED AT 15:58:10 ON 21 APR 2004  
 USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.  
 PLEASE SEE "HELP USAGETERMS" FOR DETAILS.  
 COPYRIGHT (C) 2004 American Chemical Society (ACS)

Property values tagged with IC are from the ZIC/VINITI data file  
 provided by InfoChem.

STRUCTURE FILE UPDATES: 19 APR 2004 HIGHEST RN 676225-08-4  
 DICTIONARY FILE UPDATES: 19 APR 2004 HIGHEST RN 676225-08-4

TS/CA INFORMATION NOW CURRENT THROUGH JANUARY 6, 2004

Please note that search-term pricing does apply when  
conducting SmartSELECT searches.

Crossover limits have been increased. See HELP CROSSOVER for details.

Experimental and calculated property data are now available. For more  
information enter HELP PROP at an arrow prompt in the file or refer  
to the file summary sheet on the web at:  
<http://www.cas.org/ONLINE/DBSS/registryss.html>

=>  
=>  
=> d .seq 168 1-4

L68 ANSWER 1 OF 4 REGISTRY COPYRIGHT 2004 ACS on STN  
RN 444196-48-9 REGISTRY  
CN L-Ornithine, L-cysteinyl-L-alanyl-L-seryl-L-.alpha.-glutamyl-L-asparaginyl-  
L-leucyl-L-tyrosyl-L-phenylalanyl-L-glutaminyglycyl-L-prolyl-N5-[4-  
[(iodoacetyl)amino]butyl]- (9CI) (CA INDEX NAME)  
NTE modified (modifications unspecified)

type	location	description
uncommon	Orn-12	-
modification	Orn-12	undetermined modification

SQL 12

SEQ 1 CASENLYFQG PX  
=====

HITS AT: 1-10

REFERENCE 1: 137:137275

L68 ANSWER 2 OF 4 REGISTRY COPYRIGHT 2004 ACS on STN  
RN 444196-47-8 REGISTRY  
CN L-Lysine, L-cysteinyl-L-alanyl-L-seryl-L-.alpha.-glutamyl-L-asparaginyl-L-  
leucyl-L-tyrosyl-L-phenylalanyl-L-glutaminyglycyl-L-prolyl-N6-[4-  
[(iodoacetyl)amino]butyl]- (9CI) (CA INDEX NAME)  
NTE modified (modifications unspecified)

type	location	description
modification	Lys-12	undetermined modification

SQL 12

SEQ 1 CASENLYFQG PK  
=====

HITS AT: 1-10

REFERENCE 1: 137:137275

L68 ANSWER 3 OF 4 REGISTRY COPYRIGHT 2004 ACS on STN  
RN 444196-46-7 REGISTRY  
CN L-Ornithine, L-cysteinyl-L-alanyl-L-seryl-L-.alpha.-glutamyl-L-asparaginyl-  
L-leucyl-L-tyrosyl-L-phenylalanyl-L-glutaminyglycyl-N5-[3-  
[(iodoacetyl)amino]propyl]- (9CI) (CA INDEX NAME)  
NTE modified (modifications unspecified)

type	----- location -----	description
uncommon	Orn-11 -	-
modification	Orn-11 -	undetermined modification

SQL 11

SEQ 1 CASENLYFQG X

=====

HITS AT: 1-10

REFERENCE 1: 137:137275

L68 ANSWER 4 OF 4 REGISTRY COPYRIGHT 2004 ACS on STN

RN 444196-45-6 REGISTRY

CN L-Lysine, L-cysteinyl-L-alanyl-L-seryl-L-.alpha.-glutamyl-L-asparaginyl-L-leucyl-L-tyrosyl-L-phenylalanyl-L-glutaminyglycyl-N6-[4-[(iodoacetyl)amino]butyl]- (9CI) (CA INDEX NAME)

NTE modified (modifications unspecified)

type	----- location -----	description
modification	Lys-11 -	undetermined modification

SQL 11

SEQ 1 CASENLYFQG K

=====

HITS AT: 1-10

REFERENCE 1: 137:137275

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 21, 2004, 16:15:54 ; Search time 39 Seconds  
(without alignments)  
80.902 Million cell updates/sec

Title: US-10-057-789-41\_COPY\_1\_10  
Perfect score: 56  
Sequence: 1 CASENLYFQG 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query					
No.	Score	Match	Length	DB	ID	Description
-----						

1	41	73.2	292	2	Q9X3R8	Q9x3r8 riemerella
2	41	73.2	292	2	Q93T69	Q93t69 riemerella
3	39	69.6	1615	5	Q86HW8	Q86hw8 dictyosteli
4	38	67.9	314	11	Q8VFR3	Q8vfr3 mus musculu
5	38	67.9	323	5	Q8MSJ3	Q8msj3 drosophila
6	38	67.9	372	5	Q9U1I6	Q9uli6 drosophila
7	38	67.9	372	5	Q9VII7	Q9vii7 drosophila
8	38	67.9	425	16	Q887F0	Q887f0 pseudomonas
9	38	67.9	503	10	Q9SUV4	Q9suv4 arabidopsis
10	38	67.9	531	16	Q8EZV0	Q8ezv0 leptospira
11	38	67.9	738	10	Q9LMN6	Q9lmn6 arabidopsis
12	38	67.9	738	10	O81819	O81819 arabidopsis
13	38	67.9	1366	5	Q8IQH0	Q8iqh0 drosophila
14	37	66.1	326	2	Q9X6Y4	Q9x6y4 bacteroides
15	37	66.1	377	16	Q8KDA2	Q8kda2 chlorobium
16	37	66.1	394	2	O32386	O32386 bacillus sp
17	37	66.1	415	16	Q8NU64	Q8nu64 corynebacte
18	37	66.1	499	16	Q8FLL9	Q8fll9 corynebacte
19	37	66.1	545	5	Q9V7Y2	Q9v7y2 drosophila
20	37	66.1	759	10	Q948R6	Q948r6 luffa cylin
21	36.5	65.2	439	16	Q8G870	Q8g870 bifidobacte
22	36	64.3	290	16	Q8ZMI7	Q8zmi7 salmonella
23	36	64.3	290	16	Q8Z4B5	Q8z4b5 salmonella
24	36	64.3	430	13	Q804X0	Q804x0 fugu rubrip
25	36	64.3	479	10	Q8W1N6	Q8w1n6 oryza sativ
26	36	64.3	491	6	Q863A3	Q863a3 macaca fasc
27	36	64.3	491	17	Q9HLU9	Q9hlu9 thermoplasm
28	36	64.3	518	5	Q968Y8	Q968y8 caenorhabdi
29	36	64.3	535	5	Q968Y7	Q968y7 caenorhabdi
30	36	64.3	574	10	Q8S7F4	Q8s7f4 oryza sativ
31	36	64.3	585	10	Q7XIF5	Q7xif5 oryza sativ
32	36	64.3	902	13	Q8UWC5	Q8uwc5 gallus gall
33	35	62.5	321	16	Q813B9	Q813b9 bacillus ce
34	35	62.5	420	16	Q8XMD3	Q8xmd3 clostridium
35	35	62.5	420	16	Q893E4	Q893e4 clostridium
36	35	62.5	729	3	Q8NIZ2	Q8niz2 neurospora
37	35	62.5	996	10	O24436	O24436 oryza longi
38	34	60.7	130	12	Q9J8W2	Q9j8w2 human coxa
39	34	60.7	151	4	Q9H758	Q9h758 homo sapien
40	34	60.7	152	16	Q83DC5	Q83dc5 coxiella bu
41	34	60.7	175	5	Q21315	Q21315 caenorhabdi
42	34	60.7	198	5	Q8MKK6	Q8mkk6 drosophila
43	34	60.7	219	12	O71106	O71106 bovine aden
44	34	60.7	271	2	Q9AH31	Q9ah31 pseudomonas
45	34	60.7	271	2	O34137	O34137 pseudomonas

# ALIGNMENTS

## RESULT 1

Q9X3R8

ID Q9X3R8 PRELIMINARY; PRT; 292 AA.

AC Q9X3R8;

DT 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)



DE Transposase.  
 OS *Riemerella anatipestifer*.  
 OG Plasmid pCFC2.  
 OC Bacteria; Bacteroidetes; Flavobacteria; Flavobacteriales;  
 OC Flavobacteriaceae; *Riemerella*.  
 OX NCBI\_TaxID=34085;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=20;  
 RX MEDLINE=99412224; PubMed=10481080;  
 RA Weng S., Lin W., Chang Y., Chang C.;  
 RT "Identification of a virulence-associated protein homolog gene and  
 RT ISRal in a plasmid of *Riemerella anatipestifer*."  
 RL FEMS Microbiol. Lett. 179:11-19(1999).  
 DR EMBL; AF082180; AAD33096.1; -.  
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.  
 DR InterPro; IPR002559; Transposase\_11.  
 DR Pfam; PF01609; Transposase\_11; 1.  
 KW Plasmid.  
 SQ SEQUENCE 292 AA; 34458 MW; CC81617CDF48B771 CRC64;

Query Match 73.2%; Score 41; DB 2; Length 292;  
 Best Local Similarity 70.0%; Pred. No. 6.4;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CASENLYFQG 10  
 |||: ||| |  
 Db 136 CASQKLYFYG 145

Search completed: April 21, 2004, 16:40:55  
 Job time : 50 secs

OM protein - protein search, using sw model

Run on: April 21, 2004, 16:14:59 ; Search time 12 Seconds  
 (without alignments)  
 43.392 Million cell updates/sec

Title: US-10-057-789-41\_COPY\_1\_10  
 Perfect score: 56  
 Sequence: 1 CASENLYFQG 10

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	%		DB	ID	Description
	Score	Query Match Length			
1	38	67.9	460	1 MURC_THETN	Q8r749 thermoanaer
2	38	67.9	1284	1 NRX4_DROME	Q94887 drosophila
3	37	66.1	545	1 SGPL_DROME	Q9v7y2 drosophila
4	37	66.1	1003	1 ATC_ARTSF	P35316 artemia san
5	36	64.3	47	1 YB92_HAEIN	P44125 haemophilus
6	36	64.3	348	1 YZ17_AQUAE	O66408 aquifex aeo
7	34	60.7	210	1 HAN2_XENLA	P57101 xenopus lae
8	34	60.7	309	1 HEM3_AGR5	Q8uc46 agrobacteri
9	34	60.7	448	1 XANA_XANCP	P29955 xanthomonas
10	34	60.7	451	1 G64E_DROME	P83296 drosophila
11	34	60.7	492	1 MTH8_DROME	Q9w0v7 drosophila
12	34	60.7	708	1 MM09_RAT	P50282 rattus norv
13	34	60.7	2014	1 YJU7_YEAST	P39526 saccharomyc
14	34	60.7	3054	1 POLG_TEV	P04517 t genome po
15	33	58.9	144	1 HV43_MOUSE	P01819 mus musculu
16	33	58.9	157	1 YWMA_BACSU	P70958 bacillus su
17	33	58.9	197	1 HAN1_XENLA	O73615 xenopus lae

18	33	58.9	336	1	PLSX_PSEPK	Q88118	pseudomonas
19	33	58.9	341	1	OMPU_VIBCH	P97085	vibrio chol
20	33	58.9	375	1	FMOD_BOVIN	P13605	bos taurus
21	33	58.9	376	1	FMOD_HUMAN	Q06828	homo sapien
22	33	58.9	376	1	FMOD_MOUSE	P50608	mus musculu
23	33	58.9	376	1	FMOD_RAT	P50609	rattus norv
24	33	58.9	380	1	FMOD_CHICK	P51887	gallus gall
25	33	58.9	612	1	GIDA_MYCGE	P47619	mycoplasma
26	33	58.9	626	1	MAG_MOUSE	P20917	mus musculu
27	33	58.9	626	1	MAG_RAT	P07722	rattus norv
28	33	58.9	629	1	GIDA_THEMA	Q9wya1	thermotoga
29	33	58.9	665	1	PDI2_HUMAN	Q9y2j8	homo sapien
30	33	58.9	759	1	CAS1_ARATH	P38605	arabidopsis
31	33	58.9	770	1	YRN9_CAEEL	Q09609	caenorhabdi
32	33	58.9	830	1	FAR1_YEAST	P21268	saccharomyc
33	33	58.9	868	1	MCE_ASFB7	P32094	african swi
34	33	58.9	1282	1	DOME_DROME	Q9vwe0	drosophila
35	32.5	58.0	1451	1	DPOA_RAT	O89042	rattus norv
36	32.5	58.0	1465	1	DPOA_MOUSE	P33609	mus musculu
37	32	57.1	35	1	LEC3_ULEEU	P23032	ulex europe
38	32	57.1	200	1	PIGB_ELAAU	Q9pwi3	elaphe quad
39	32	57.1	290	1	HYPB_ECOLI	P24190	escherichia
40	32	57.1	294	1	ATHB_RAT	P18598	rattus norv
41	32	57.1	413	1	HUTI_FUSNN	Q8rfgl	fusobacteri
42	32	57.1	419	1	CCA_BUCBP	Q89b06	buchnera ap
43	32	57.1	440	1	Y48K_ELV	P35929	erysimum la
44	32	57.1	444	1	GID_STRMU	P05428	streptococc
45	32	57.1	522	1	GUAÄ_WIGBR	Q8dlv0	wiggleswort

# ALIGNMENTS

## RESULT 1

### MURC\_THETN

ID MURC THETN STANDARD; PRT; 460 AA.  
AC Q8R749;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE UDP-N-acetylmuramate--L-alanine ligase (EC 6.3.2.8) (UDP-N-  
DE acetylmuramoyl-L-alanine synthetase).  
GN MURC OR TTE2575.  
OS Thermoanaerobacter tengcongensis.  
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;  
OC Thermoanaerobacteriaceae; Thermoanaerobacter.  
OX NCBI\_TaxID=119072;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MB4 / JCM 11007;  
RX MEDLINE=21992816; PubMed=11997336;  
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,  
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,  
RA Tan H., Chen R., Wang J., Yu J., Yang H.;  
RT "A complete sequence of T. tengcongensis genome."  
RL Genome Res. 12:689-700(2002).  
CC -!- FUNCTION: Cell wall formation.

```

CC  -!- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmuramoyl + L-alanine = ADP +
CC      phosphate + UDP-N-acetylmuramoyl-L-alanine.
CC  -!- PATHWAY: Peptidoglycan biosynthesis.
CC  -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC  -!- SIMILARITY: Belongs to the murCDEF family.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AE013198; AAM25699.1; -.
DR  HAMAP; MF_00046; -; 1.
DR  InterPro; IPR000713; Mur_ligase.
DR  InterPro; IPR004101; Mur_ligase_C.
DR  InterPro; IPR005758; MurC.
DR  Pfam; PF01225; Mur_ligase; 1.
DR  Pfam; PF02875; Mur_ligase_C; 1.
DR  TIGRFAMs; TIGR01082; murC; 1.
KW  Ligase; ATP-binding; Cell division; Cell wall;
KW  Peptidoglycan synthesis; Complete proteome.
FT  NP_BIND      116      122      ATP (POTENTIAL).
SQ  SEQUENCE      460 AA;  51282 MW;  5C2C1724CF7BD71B CRC64;

```

```

Query Match          67.9%;  Score 38;  DB 1;  Length 460;
Best Local Similarity 60.0%;  Pred. No. 6.4;
Matches      6;  Conservative      3;  Mismatches      1;  Indels      0;  Gaps      0;

```

```

Qy          1 CASENLYFQG 10
            ||| |:|::|
Db          259 CASFNVYYRG 268

```

```

Search completed: April 21, 2004, 16:40:01
Job time : 21 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 21, 2004, 16:39:36 ; Search time 21 Seconds  
(without alignments)  
45.805 Million cell updates/sec

Title: US-10-057-789-41\_COPY\_1\_10  
Perfect score: 56  
Sequence: 1 CASENLYFQG 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_78:\*  
1: pirl:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	% Query		Match	Length	DB	ID	Description
	Score	Match					
1	38	67.9	503	2	T05347	hypothetical prote	
2	38	67.9	738	2	D86345	hypothetical prote	
3	38	67.9	1283	2	T13799	neurexin IV - frui	
4	37	66.1	674	2	S32230	Ca2+-transporting	
5	37	66.1	1003	2	S07526	Ca2+-transporting	
6	36	64.3	47	2	H64021	hypothetical prote	
7	36	64.3	290	2	AC0847	hydrogenase isoenz	
8	36	64.3	419	2	T32441	hypothetical prote	
9	35	62.5	996	2	T10725	protein kinase Xa2	
10	34	60.7	175	2	T23437	hypothetical prote	
11	34	60.7	283	2	T19732	hypothetical prote	
12	34	60.7	304	2	AD0864	probable membrane	
13	34	60.7	309	2	AC2902	porphobilinogen de	

14	34	60.7	309	2	E97677	porphobilinogen de
15	34	60.7	448	2	A43304	phosphomannomutase
16	34	60.7	462	2	AH1053	probable exported
17	34	60.7	500	2	G82828	phosphoglucomutase
18	34	60.7	708	2	JC4364	gelatinase B (EC 3
19	34	60.7	708	2	S62907	gelatinase B (EC 3
20	34	60.7	724	2	T00495	L5 protein - white
21	34	60.7	908	2	AE2675	pyruvate, orthophos
22	34	60.7	933	2	C97457	pyruvate, phosphat
23	34	60.7	2014	2	S46622	probable membrane
24	34	60.7	3054	1	GNBVEV	genome polyprotein
25	33	58.9	144	1	G2MS14	Ig heavy chain pre
26	33	58.9	157	2	E70062	hypothetical prote
27	33	58.9	177	2	AH2393	hypothetical prote
28	33	58.9	215	2	F96746	probable drought i
29	33	58.9	283	2	T22567	hypothetical prote
30	33	58.9	287	2	A81803	probable integral
31	33	58.9	288	2	D81065	hypothetical prote
32	33	58.9	350	2	A82299	outer membrane pro
33	33	58.9	355	2	T08703	alpha-catenin homo
34	33	58.9	368	2	C71975	hypothetical prote
35	33	58.9	368	2	D64532	conserved hypothet
36	33	58.9	375	2	S05390	fibromodulin precu
37	33	58.9	376	2	S55275	fibromodulin precu
38	33	58.9	380	2	S71876	fibromodulin - chi
39	33	58.9	382	2	T01162	hypothetical prote
40	33	58.9	435	2	A82554	conserved hypothet
41	33	58.9	475	2	D88451	protein K10D2.2 [i
42	33	58.9	582	1	BNRT3S	myelin-associated
43	33	58.9	612	2	I64241	glucose inhibited
44	33	58.9	626	1	BNRT3	myelin-associated
45	33	58.9	629	2	F72400	glucose-inhibited

# ALIGNMENTS

## RESULT 1

T05347

hypothetical protein F8B4.70 - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 23-Jul-1999

C;Accession: T05347

R;Bevan, M.; Terry, N.; Ardiles, W.; Buysschaert, C.; Dasseville, R.; De Clerck, R.; De Keyser, A.; Neyt, P.; Rouze, P.; Van Den Daele, H.; Villaroel, R.; Gielen, J.; Van Montagu, M.; Hoheisel, J.; Mewes, H.W.; Mayer, K.F.X.; Schueller, C.

submitted to the Protein Sequence Database, February 1999

A;Reference number: Z15409

A;Accession: T05347

A;Molecule type: DNA

A;Residues: 1-503 <BEV>

A;Cross-references: EMBL:AL034567

A;Experimental source: cultivar Columbia; BAC clone F8B4

C;Genetics:

A;Map position: 4

A;Introns: 65/3; 109/3; 170/3; 213/3; 261/3; 300/3

A;Note: F8B4.70

Query Match 67.9%; Score 38; DB 2; Length 503;  
Best Local Similarity 60.0%; Pred. No. 13;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CASENLYFQG 10  
||| |:: ||  
Db 253 CASRNIHIQG 262

Search completed: April 21, 2004, 16:54:40  
Job time : 24 secs

OM protein - protein search, using sw model

Run on: April 21, 2004, 16:14:24 ; Search time 56 Seconds  
 (without alignments)  
 50.455 Million cell updates/sec

Title: US-10-057-789-41\_COPY\_1\_10  
 Perfect score: 56  
 Sequence: 1 CASENLYFQG 10

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
 1: geneseqp1980s:\*  
 2: geneseqp1990s:\*  
 3: geneseqp2000s:\*  
 4: geneseqp2001s:\*  
 5: geneseqp2002s:\*  
 6: geneseqp2003as:\*  
 7: geneseqp2003bs:\*  
 8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	56	100.0	11	5	ABG96242	Abg96242 Peptide 1
2	56	100.0	11	5	ABG96241	Abg96241 Peptide 1
3	56	100.0	11	6	ABP95983	Abp95983 Link grou
4	56	100.0	11	6	ABP95982	Abp95982 Link grou
5	56	100.0	12	5	ABG96243	Abg96243 Peptide 1
6	56	100.0	12	5	ABG96244	Abg96244 Peptide 1
7	56	100.0	12	6	ABP95984	Abp95984 Link grou
8	56	100.0	12	6	ABP95985	Abp95985 Link grou
9	47	83.9	19	5	ABG95951	Abg95951 Peptide 1



10	47	83.9	19	5	ABG95936	Abg95936 Peptide 1
11	47	83.9	19	5	ABG95984	Abg95984 Peptide e
12	47	83.9	19	5	ABG95961	Abg95961 Peptide 1
13	47	83.9	19	5	ABG95972	Abg95972 Peptide e
14	47	83.9	19	5	ABG96240	Abg96240 Peptide e
15	47	83.9	19	5	ABG95971	Abg95971 Peptide e
16	47	83.9	19	5	ABG95941	Abg95941 Peptide 1
17	47	83.9	19	5	ABG95946	Abg95946 Peptide 1
18	47	83.9	19	5	ABG95956	Abg95956 Peptide 1
19	47	83.9	19	6	ABP95945	Abp95945 Link grou
20	47	83.9	19	6	ABP95950	Abp95950 Link grou
21	47	83.9	19	6	ABP95980	Abp95980 Reagent p
22	47	83.9	19	6	ABP95970	Abp95970 Reagent p
23	47	83.9	19	6	ABP95965	Abp95965 Reagent p
24	47	83.9	19	6	ABP95955	Abp95955 Link grou
25	47	83.9	19	6	ABP95979	Abp95979 Reagent p
26	47	83.9	19	6	ABP95960	Abp95960 Reagent p
27	47	83.9	20	5	ABG95949	Abg95949 Peptide 1
28	47	83.9	20	5	ABG95937	Abg95937 Peptide 1
29	47	83.9	20	5	ABG95945	Abg95945 Peptide 1
30	47	83.9	20	5	ABG95952	Abg95952 Peptide 1
31	47	83.9	20	5	ABG95965	Abg95965 Peptide 1
32	47	83.9	20	5	ABG95938	Abg95938 Peptide 1
33	47	83.9	20	5	ABG95947	Abg95947 Peptide 1
34	47	83.9	20	5	ABG95954	Abg95954 Peptide 1
35	47	83.9	20	5	ABG95955	Abg95955 Peptide 1
36	47	83.9	20	5	ABG95959	Abg95959 Peptide 1
37	47	83.9	20	5	ABG95939	Abg95939 Peptide 1
38	47	83.9	20	5	ABG95942	Abg95942 Peptide 1
39	47	83.9	20	5	ABG95943	Abg95943 Peptide 1
40	47	83.9	20	5	ABG95948	Abg95948 Peptide 1
41	47	83.9	20	5	ABG95953	Abg95953 Peptide 1
42	47	83.9	20	5	ABG95964	Abg95964 Peptide 1
43	47	83.9	20	5	ABG95957	Abg95957 Peptide 1
44	47	83.9	20	5	ABG95962	Abg95962 Peptide 1
45	47	83.9	20	5	ABG95940	Abg95940 Peptide 1

# ALIGNMENTS

## RESULT 1

ABG96242

ID ABG96242 standard; peptide; 11 AA.

XX

AC ABG96242;

XX

DT 11-DEC-2002 (first entry)

XX

DE Peptide link group, used to isolate functional groups in proteins, #36.

XX

KW Rabbit; bovine; analytical reagent; trifunctional; peptide mixture;

KW enrichment; immobilisation site; cleavage site; link; epitope tag;

KW protease; cysteine-containing; perturbed cell; mass spectrometry;

KW peptide tag; BSA; bovine serum albumin; PEPTag; APEPTag; IPEPTag;

KW affinity peptide encoded tag; immobilised peptide encoded tag; chicken;

KW beta-lactoglobulin; GAPDH; glyceraldehyde-3-phosphate dehydrogenase;

KW a-lactalbumin; ovalbumin; yeast.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200259144-A2.  
 XX  
 PD 01-AUG-2002.  
 XX  
 PF 25-JAN-2002; 2002WO-US002487.  
 XX  
 PR 26-JAN-2001; 2001US-0264576P.  
 PR 13-JUL-2001; 2001US-0305232P.  
 XX  
 PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
 XX  
 PI Haynes P, Wei J, Yates J, Andon N;  
 XX  
 DR WPI; 2002-599760/64.  
 XX  
 PT Novel trifunctional synthetic reagents for labeling peptides at specific  
 PT amino acid residue and selectively enriching only those peptides  
 PT containing labeled amino acid, useful for proteomic analysis.  
 XX  
 PS Claim 33; Page 57; 79pp; English.  
 XX  
 CC The invention discloses analytical reagents (e.g. trifunctional synthetic  
 CC reagents) which can be used for reducing the complexity of peptide  
 CC mixtures. The method labels peptides at a specific amino acid residue and  
 CC then selectively enriches only those peptides containing the labelled  
 CC amino acid. The compound have the formula of immobilisation site-cleavage  
 CC site-link. The immobilisation site is chosen from an epitope tag, a  
 CC linker to a solid surface, a metal chelating site, a magnetic site and a  
 CC specific oligonucleotide sequence, or their combination, the cleavage  
 CC site is chosen from a protease cleavage site, a photocleavable linker, a  
 CC restriction enzyme cleavage site, a chemical cleavage site and a thermal  
 CC cleavage site, or their combination and the link is chosen from an amino  
 CC acid reactive site and a mass variance site, or their combination. The  
 CC compounds are useful for simultaneously identifying and determining the  
 CC levels of expression of cysteine-containing proteins in normal and  
 CC perturbed cells. The advantage is that these reagents allow rapid and  
 CC quantitative analysis of proteins or protein function in mixtures of  
 CC proteins. By preparing the reagent in two forms with detectably different  
 CC masses, accurate relative quantification of peptide amounts using mass  
 CC spectrometry, can be achieved. The sequences given in ABG95935-ABG96244  
 CC are examples of the peptide tags used to isolate cysteine-containing  
 CC proteins, the target sequences tested and the peptides isolated using the  
 CC peptide tags  
 XX  
 SQ Sequence 11 AA;

Query Match 100.0%; Score 56; DB 5; Length 11;  
 Best Local Similarity 100.0%; Pred. No. 0.0004;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CASENLYFQG 10  
 |||||  
 Db 1 CASENLYFQG 10

RESULT 3

ABP95983

ID ABP95983 standard; peptide; 11 AA.

XX

AC ABP95983;

XX

DT 30-APR-2003 (first entry)

XX

DE Link group related peptide #2.

XX

KW Protease cleavage site; proteomic difference; peptide expression;  
KW analysis; qualitative; quantitative; detection; identification.

XX

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Modified-site 1

FT /note= "acetylated"

FT Modified-site 11

FT /label= Orn

FT /note= "Ornithine is C-terminally attached to CH2CH2CH2-

FT NH-C(O)-CH2I"

XX

PN WO2003006951-A2.

XX

PD 23-JAN-2003.

XX

PF 12-JUL-2002; 2002WO-US022320.

XX

PR 13-JUL-2001; 2001US-0305169P.

PR 21-FEB-2002; 2002US-0359524P.

XX

PA (SYGN ) SYNGENTA PARTICIPATIONS AG.

XX

PI Washburn M, Deciu C, Ulasek R;

XX

DR WPI; 2003-221772/21.

XX

PT System for detecting peptide expression levels, comprises a mixture of 2  
PT labeled peptides from 2 samples and modules to calculate peptide weight,  
PT identify peptide pair and quantify abundance of each peptide in the pair.

XX

PS Disclosure; Page 54; 110pp; English.

XX

CC The present invention describes a system (I) for detecting peptide  
CC expression levels, which comprises a peptide mixture (PM) having first  
CC and second labeled peptides from first and second biological samples  
CC (S1,S2), respectively (peptides having same sequence in S1, S2 have  
CC predetermined mass difference), and three modules to calculate peptide  
CC weight in PM, to identify peptide pair (PP) in PM, and to quantify  
CC abundance of each peptide in PP. Also described: (1) a method (M) for  
CC detecting peptide expression levels between a first biological sample and  
CC a second biological sample; and (2) a system (II) for proteomic analysis

CC of two or more peptide populations. (I), (II) or (M) is useful for  
CC detecting and quantifying proteomic differences between two or more  
CC biological samples. (I) is useful for the quantitative analysis of  
CC peptide expression in complex samples (such as cells, tissues and their  
CC fractions), for differential expression analysis between multiple samples  
CC and the identification of novel peptides, or for simultaneously  
CC identifying and detecting the expression levels of cysteine-containing  
CC protein in normal and perturbed cells. (M) is useful for qualitative and  
CC quantitative analysis of global expression profiles in cells and tissues,  
CC i.e., the quantitative analysis of proteomes, to implement a variety of  
CC clinical and diagnostic analyses to detect the presence, absence,  
CC deficiency or excess of a given protein or protein function in a  
CC biological fluid (e.g., blood), or in cells or tissues, or for analysis  
CC of complex mixture of proteins. The present sequence represents a Link  
CC group peptide which is given in the exemplification of the present  
CC invention

XX

SQ Sequence 11 AA;

Query Match 100.0%; Score 56; DB 6; Length 11;

Best Local Similarity 100.0%; Pred. No. 0.0004;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CASENLYFQG 10

|||||||

Db 1 CASENLYFQG 10

Search completed: April 21, 2004, 16:39:28

Job time : 57 secs